UNIVERSITÉ DE MONTRÉAL

CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

CRT-584

EQUILIBRIUM ASSIGNMENT ON CONGESTED TRANSIT NETWORKS

by

Joaquin DeCea*
and
Enrique Fernandez**

* Pontificia Universidad Catolica de Chile, Santiago, Chile.

** Pontificia Universidad Catolica de Chile, Santiago, Chile and Visiting Professor at the Centre de recherche sur les transports, Université de Montréal, Montréal, Canada.

This research has been financed by grants from the Research Division of the Pontificia Universidad Catolica de Chile (DIUC), the Chilean National Fund for Scientific and Technological Development (FONDECYT), and the International Development Research Centre (IDRC).

Cette étude a été publiée grâce à une subvention du fonds F.C.A.R. pour l'aide et le soutien à la recherche.

Université de Montréal
Centre de recherche sur les transports - Publication #584
July 1988

Dépôt légal - 3e trimestre 1988

ABSTRACT

We propose a new formulation for the assignment problem over congested transit networks. The congestion effects due to insufficient capacity of system elements (transit lines) are considered to be concentrated at transit stops. Waiting times on access links are therefore dependent on passenger flows. A special formulation of the transit network is used in order to model correctly the congestion effects. Finally, algorithms for solution are analyzed.

Key words : transit networks, equilibrium, assignment, congestion.

- - - - -

RÉSUMÉ

Nous proposons une nouvelle formulation pour le problème d'affectation des passagers dans un réseau de transport en commun congestionné. Les effets de congestion résultant d'une sous-capacité du système (fréquences trop basses) sont pris en compte aux arrêts. Ainsi les temps d'attente sur les liens d'accès sont dépendants des flots de passagers. Une nouvelle représentation du réseau de transport en commun permet de modéliser adéquatement les effets de congestion. Nous proposons également des algorithmes de résolution pour le modèle.

Mots-clés : réseau de transport en commun, affectation, équilibre, congestion.

1. Introduction.

One of the main limitations of existing transit assignment models is that they do not consider congestion effects over the transit system. The assumption is made that all transit lines have unlimited capacity to accommodate any amount of demand that they could face. Such models with passengers travel times independent of network flows, are formulated as linear programs for which efficient solution algorithms have been proposed (see Spiess and Florian, 1988; De Cea et al., 1988) .

Although the existing models can be useful to study some transit systems where the congestion effects are not important (estimation of transit flows on transit networks of developed countries), they present important limitations for the study of transit networks in developing countries, where high congestion levels are usually observed due to insufficient capacity of the transit system. In addition, linear models can not be properly used as part of more complex models such as transit network design models and operators equilibrium models, given that they produce highly unrealistic results (see Fernández and Bravo, 1986).

Some attempts have been made in the past to build transit assignment models that consider congestion, however the way they represent congestion is limited and does not take into account the main phenomena as occurring in practice.

TRANSEPT (Last and Leak, 1976) uses an iterative process in order to assign trips to common lines considering capacity and waiting times. Nevertheless, the trip loading process is performed link by link for each route, which makes the model appropriate only for radial networks.

Spiess (1983) gives a general version of his linear model in which in-vehicle travel times (or generalized travel costs) are increasing functions of passenger flows (the author calls them "discomfort function"). However, as the same author acknowledges, the model presents important limitations, the main one being that

waiting times at stops are not affected by transit volumes, reducing the congestion phenomenon to a comfort problem.

A similar formulation is proposed by Nguyen and Pallotino (1988). They introduce the concept of hyperpath and formulate an equilibrium model similar to that used for congested road networks. Nevertheless as in the Spiess model, waiting times are considered constant and independent of trip volumes.

In this work we present a new model formulation for the transit assignment problem on congested systems. It relates congestion to the waiting phenomena experienced by travellers at stops when the trip volumes approach the system capacity. In section 2 some basic definitions and the notation are introduced. Then, section 3 presents and discusses the main modelling assumptions. Section 4 deals with a network transformation to properly represent the congestion phenomenon on a transit network. In section 5 we give an analytical formulation of the transit assignment problem and finally in section 6 we discuss some solution approaches. In the Appendix we present some small examples to illustrate the application of the model and compare its results with those provided by models currently in use.

2. Basic Definitions and Notation.

2.1 Definitions.

Given that we will be frequently using the terms line, itinerary, route, line section and route section, we will start defining what we mean by each of them.

A "transit line" or just a "line", is a group of vehicles that run back and forth between two nodes on the urban road network. All vehicles in the same line have identical sizes, capacities and operating characteristics and travel on the network going always through the same sequence of network links and nodes, referred to as an "itinerary".[1] We assume that each vehicle stops at each node included

[1] We can also refer to the "itinerary" of a particular trip, which then corre-

in its itinerary, in order to allow passengers to board or alight; therefore, those road network nodes where vehicles do not stop, do not need to be included in the description of an itinerary. Each transit line provides a public transportation service defined by its itinerary, and the vehicles frequency and capacity.

A "line section" is any portion of a transit line between two non necessarily consecutive nodes of its itinerary.

A "transit route" is any path that a transit user can follow on the transit network in order to travel between any two nodes. In general it will be identified by a sequence of nodes, the first one being the origin of the trip, the final being the destination and all the intermediate nodes represent transfer points. The portion of a route between two consecutive transfer nodes is called "route section". Each route section is associated with a set of "attractive" or "common" lines.

2.2 Notation.

Consider a network $G = (\mathcal{N}, \mathcal{S})$, with node set $\mathcal{N}$, and link set $\mathcal{S}$. We will denote:

W : set of network origin-destination (O-D) pairs.

w : an element of set W.

P : set of paths in G available to transit users.

P_w : set of feasible paths associated with O-D pair w.

p : an element of set P.

T : set of O-D transit demands.

T_w : transit demand between O-D pair w.

V^s : transit passenger flow on link s.

h_p : transit passenger flow over path p.

---

sponds to the sequence of line sections used by a particular traveler between a specific O–D pair.

c_s : travel cost for transit users, on link s.

$\bar{c}_s$: in vehicle travel time on link s.

C_p : travel cost for transit users, over path p.

$\{\delta_{sp}\}$: transit link-path incidence matrix:

-takes a value of 1 if arc a belongs to path p,

-and 0 otherwise.

For the transit lines we will use:

v_l^s : passenger flow on line l, section s.

f_l : frequency of line l.

k_l : practical capacity of line l.

f_s : total frequency on link s.

K_s : practical capacity on link s.

3. Transit system operation and modelling assumptions.

In general there will be many different transit routes that a passenger could use to travel between two nodes on the transit network. We here assume that he selects that transit route that minimizes his total travel time (in-vehicle plus waiting).

Given the way that we defined transit routes, if we also assume that travel times are independent of the flows using the transit services, for each given O-D pair passengers should be assigned to the transit route that provides the minimum travel time, obtaining the same solution provided by Chriqui's algorithm (Chriqui, 1974).

However, we will assume that the transit system has a limited capacity and therefore, as passenger flows increase, travel times also increase. Then, as some routes become congested passengers will consider using alternative routes that become attractive. As congestion increases so will the number of alterna-

tive routes considered. This is similar to what is usually assumed for modelling congested road networks but as we will see later a somewhat different network representation will be used for the transit case.

We will assume that the congestion phenomena on transit networks is concentrated at transit stops. There, passengers experience waiting times that depend on the total capacity of the attractive set of lines considered and on the total number of passengers using that set of lines. We can look at this situation as a queueing process: as the number of passengers trying to use a given service approaches its capacity, waiting times increase. We will assume however that once a passenger boards a line the in-vehicle travel time is fixed and solely determined by the level of congestion over the road network, which will be considered an exogenous parameter. This is an important difference with other formulations proposed for congested transit networks, which assume that congestion affects in-vehicle travel times but not waiting times (Spiess, 1983; Nguyen and Pallotino, 1988). Such formulations have been justified by claiming that increasing functions for in-vehicle travel times represent the increasing discomfort experienced by passengers riding crowded vehicles.

When a passenger is traveling over a given transit route, at each transit stop (origin or transfer) he will consider a set of attractive lines and he will board the first vehicle belonging to those lines that has a place available.

Passengers waiting at the same transit stop can be classified in different groups, depending on their next transfer node, given that each of them will have a different set of attractive lines; therefore the services each group is waiting for will be different, with different capacities, utilizations and waiting times.

With respect to the definition of the set of attractive or common lines for a passenger traveling between a given pair of transfer nodes, we will assume that different sets can be defined depending on the level of congestion in the system.

The most general approach would be to consider that all lines connecting two nodes are "potentially" attractive. Thus, if we have both "fast" and "slow"[2] lines connecting two nodes, only the first ones will be considered when there is no congestion; nevertheless when flows increase to the point that they do not have enough capacity to carry everybody, then the slow lines will also become attractive.

In terms of total travel times, as the proportion of full "fast" vehicles arriving to the transit stop increases, waiting times for the fast lines will increase to the point that travelers will consider the use of "slow" lines.

4. Network Definition.

4.1 Network representation.

Given the assumptions of the previous section we will use a transit network $G = (\mathcal{N}, \mathcal{S})$ with node set $\mathcal{N}$, representing transit stops and link set $\mathcal{S}$, representing network links which in our case will correspond to:

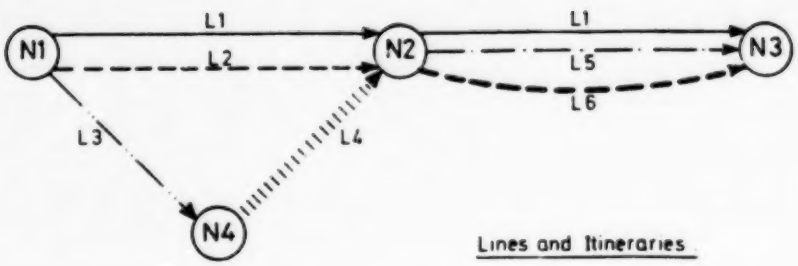
i) Waiting links, representing the waiting process experienced at transit stops.

ii) In vehicle travel links, corresponding to route sections.

Although these are the two basic types of links, nevertheless a third type could also be considered, corresponding to walk access links with a constant travel time and infinite frequency.

For each node of the transit network there will be one waiting link associated with each route section leaving the node. It will represent the average waiting time experienced at that stop by the passengers traveling on any route using that

[2] Note that fast and slow lines are defined only in terms of their in-vehicle travel times. Thus, for a given pair of nodes a fast line could be one that uses a short direct route and a slow line one that goes through a longer route passing through other intermediate nodes.



Lines and Itineraries

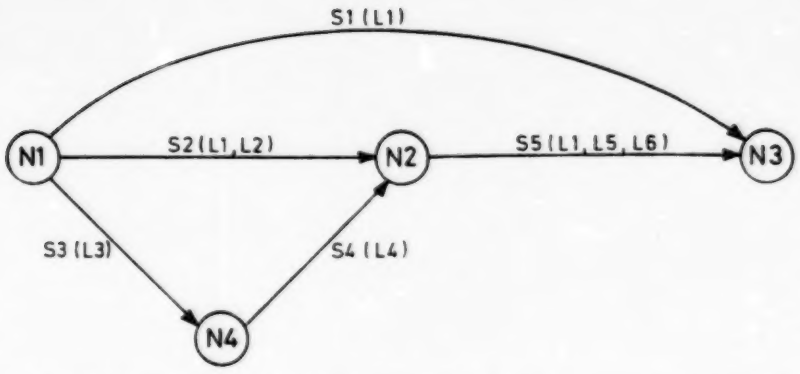
L1	:	N1 , N2 , N3
L2	:	N1 , N2
L3	:	N1 , N4
L4	:	N4 , N2
L5	:	N2 , N3
L6	:	N2 , N3

FIG. 1. Network representation using lines, and itineraries description.

section. Therefore, given the one-to-one relation between waiting links and travel links, we can make them collapse into a single link in the network representation. Then the link cost function should be appropriately defined as we will see below.

In order to codify our transit network we need to identify all the route sections and the corresponding set of attractive lines going through the pair of nodes that defines this route section. In Figure 1, a network is represented in terms of its transit lines and the itineraries are described for each line. Figure 2 represents the same network but codified in terms of route sections; routes and route sections are described for the O-D pair (N_1,N_3).

It is important to note that in Figure 2, S_2 is a route section that can be used both by passengers that travel from origin N_1 to destination N_3, using route R_2 (transferring at node N_2) and passengers going from origin N_1 to destination N_2, using S_2. Also, S_5 can be used by passengers traveling between N_1 and N_3 and passengers with origin in N_2 and destination N_3. Route section S_1 is used by passengers that travel between N_1 and N_3 directly (without transfers) and in



Routes	Transfert nodes	Routes sections
R_1	(N1, N3)	$\{(N1,N3) \xrightarrow[S1]{} L1\}$
R_2	(N1, N2, N3)	$\{(N1,N2) \xrightarrow[S2]{} (L1\ L2)\}, \{(N2,N3) \xrightarrow[S5]{} (L1, L5, L6)\}$
R_3	(N1, N4, N2, N3)	$\{(N1,N4) \xrightarrow[S3]{} L3\}, \{(N4,N2) \xrightarrow[S4]{} L4\}, \{(N2,N3) \xrightarrow[S5]{} (L1, L5, L6)\}$

FIG. 2. Network representation using route sections

this case they can only use line L_1. By introducing all the attractive lines that go through each pair of possible transfer nodes, without considering the origin and final destination of travelers going through them, we dramatically reduce the number of links (route sections) that are necessary to consider; however, this requires the use of link cost functions that properly take into account the interactions between different route sections that share the same capacity.

4.2 Link cost functions.

When there is no flow on the network, or congestion is negligible, we will assume that the time (or cost) spent to traverse a link s will be constant and equal to the in-vehicle travel time $\bar{c}_s$ plus the expected waiting time, which will be proportional to the inverse of the total frequency over link f_s. Then we will have:

$$\tilde{c}_s = \bar{c}_s + \left(\frac{\alpha}{f_s}\right) \tag{1}$$

where α is a parameter whose value depends on the distribution assumed for buses interarrival times (headways) and passenger arrival times. The normally used value $\alpha = .5$, corresponds to a uniform random variable.

Nevertheless, when flow increases over the network congestion will make waiting times to increase so that an additional term, representing the effect of congestion, must be incorporated into (1). Given that our network representation considers different links sharing the capacity of the same lines, the congestion experienced on one link will depend not only on its own flow V^s, but also on other link flows.

In general, the waiting time experienced by a passenger boarding link s (at its head node $i(s)$) will depend on:

i) V^s, the total number of passengers boarding the same link, at node is,

ii) V_{is}^+, the total number of passengers boarding, at the same node, all other links that use lines contained in link s, and,

iii) $\overline{V}_{is}$, the number of passengers boarding all the lines belonging to link s at a node before $i(s)$ and alighting after $i(s)$.

The flow V_{is}^+ competes with V^s for the residual capacity of link s lines at node $i(s)$, and the flow $\overline{V}_{is}$ reduces that capacity.

Then, calling S_{is}^+ to the set of links going out of node is, with the exception of link s, and $\overline{S}_{is}$ to the set of links with initial node before $i(s)$ and final node after $i(s)$, we can define:

$$\widetilde{V}_s = \sum_{l \in s} \underbrace{\sum_{r \in S_{is}^+} v_l^r}_{V_{is}^+} + \underbrace{\sum_{r \in \overline{S}_{is}} v_l^r}_{\overline{V}_{is}} \qquad (2)$$

where $\widetilde{V}_s$ represents the flows that compete with V^s for the same common capacity.

Then, the link cost functions (generally called volume-delay functions) will

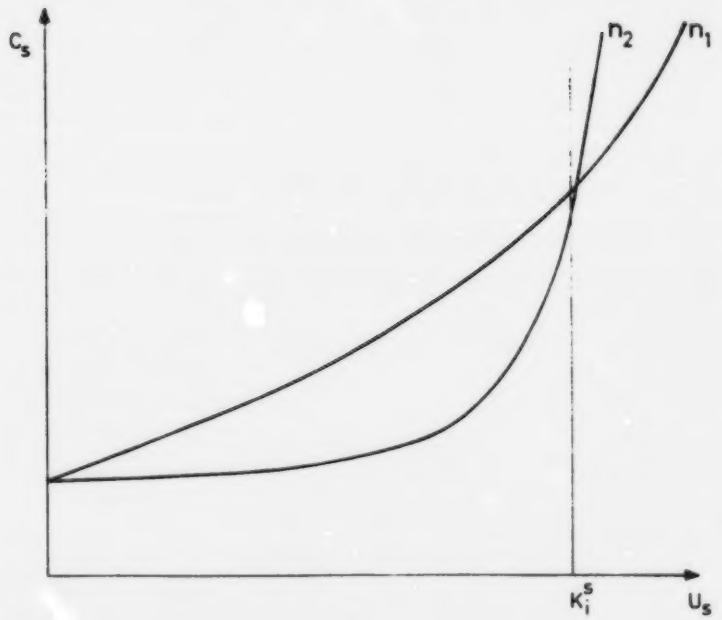


FIG. 3. Link cost functions general form.

have the following general form:

$$c_s = \bar{c}_s + \left(\frac{\alpha}{f_s}\right) + \beta \cdot \varphi_s \left(\frac{V^s + \tilde{V}_s}{K_s}\right) \tag{3}$$

The general form of function c_s is shown in Figure 3. Note that K_s is not in *strictu sense* a capacity, but something similar to the "practical capacity" used to define link cost functions in road networks. The form of φ_s should be such that c_s is strictly monotone in V^s. One possibility is the power form used in BPR functions:

$$\varphi_s(V) = \left(\frac{V^s + \tilde{V}_s}{K_s}\right)^n$$

In such case as n increases, K_s acts more and more as a "true" capacity; in Figure 3, n_1 corresponds to a small value of power n and n_2 to a larger value. No matter that the function will be the same for all links in the network, it is easy

easy to see that the Jacobian of the vector cost function c (with elements c_s) will not be in general symmetric, given that different lines have different capacities. As a consequence, the congested transit assignment problem is an asymmetric network flow problem.

5. Mathematical Formulation.

Any feasible set of flows $\{V\}$ in $G(\mathcal{N},\mathcal{S})$ satisfies conservation constraints:

$$\sum_{p\in P_w} h_p = T_w, \qquad \forall w \in W \tag{4}$$

$$\sum_{p\in P} \delta_{sp}\cdot h_p = V^s, \qquad \forall s \in S \tag{5}$$

As discussed in section 3, a general formulation should consider as attractive all lines connecting a pair of transfer nodes. Then, the set actually considered by passengers travelling between both nodes will depend on the congestion level. Therefore, in order to identify the lines included in the set $\overline{A}_s$ of attractive lines belonging to route section s, the following binary variables will be defined:

$$x_l^s = \begin{cases} 0, & \text{if } l \notin \overline{A}_s, \\ 1, & \text{if } l \in \overline{A}_s, \end{cases} \qquad \forall l \in s, \tag{6}$$

Thus, when there is no congestion, waiting time for passengers boarding link s can be expressed as:[3]

$$\overline{\omega}^s = \frac{1}{\sum_{l\in s} x_l^s \cdot f_l} \qquad \text{with:} \quad f_s = \sum_{l\in s} x_l^s \cdot f_l, \tag{7}$$

However, when congestion exists, a fraction of buses arriving at node $i(s)$ will be full. This fraction will increase as flows approach system capacity and the effective frequency for somebody waiting for a bus from section s will be less than

[3] In the following we will assume $\alpha = 1$, or expected waiting times equal to headways (Poisson arrivals)

f_s. Actually, its value will tend to zero as flows approach capacity and waiting times increase.

We will define f'_s, the effective frequency for link s as:

$$f'_s = \sum_{l \in s} x_l^{s\prime} \cdot f_l^{s\prime}, \qquad \text{where:} \quad f_l^{s\prime} = \left(\frac{1}{\omega_l^s}\right), \qquad \text{and:} \quad x_l^{s\prime} = x_l^{s\prime}(V), \tag{8}$$

with:

$$\omega_l^s \overset{\text{def}}{=} \frac{\alpha_l}{f_l} + \beta_l \cdot \left(\frac{v_l^s + \widetilde{v}_l}{k_l}\right)^n, \tag{9}$$

where v_l^s is the flow that boards line l at node is and $\widetilde{v}_l$ are all the flows competing for the capacity of line l defined similarly than in $\widetilde{V}_s$. Note that, in order to have a consistent formulation, from (3), (8) and (9) the following must hold:

$$\omega^s(V) = \frac{1}{\sum_{l \in s} x_l^{s\prime} \cdot f_l^{s\prime}} \tag{10}$$

or:

$$\varphi_s(V, K) = \left[\sum_{l \in s} \left(\frac{x_l^{s\prime} \cdot k_l}{v_l^s + \widetilde{v}_l}\right)^n\right]^{-1} \tag{11}$$

Then, the relation between route section flows and line section flows over our transit network will be given by:

$$v_l^s = x_l^{s\prime} \cdot f_l^{s\prime} \cdot \omega^s(V) \cdot V^s \tag{12}$$

Any set of flows $\{v_l^s\}$ will be said to be **feasible**, if it satisfies equations (4), (5) and (12), in addition to the following non-negativity and integrality constraints:

$$v_l^s \geq 0, \qquad x_l^{s\prime} = \{0, 1\}. \tag{13}$$

The set of feasible flows will be denoted by Ω. Equations (4) and (5) are linear as well as the non-negativity constraints. However, equations (12) are non linear and the definition of variables v_l^s includes integer variables $x_l^{s\prime}$.

A feasible flow set will be called an **equilibrium** if it satisfies the following Wardropian conditions over the transit network $G(\mathcal{N},\mathcal{S})$:

$$C_p \begin{cases} = u_w, & \forall p \in P_w \quad / \quad h_p \geq 0; \\ > u_w, & \forall p \in P_w \quad / \quad h_p = 0, \end{cases} \qquad \forall w \in W, \tag{14}$$

where u_w is the equilibrium travel cost over all used paths that connect w.

It is easy to show (see Florian and Spiess, 1983), that conditions (14) have an equivalent variational formulation. Therefore, an equilibrium can be represented by the following variational problem:

$$c(V^*) \cdot (V^* - V) \leq 0, \qquad \forall V \in \Omega, \tag{15}$$

where c is the vector of route section costs, V is any feasible vector of route section flows $\{V^s\}$, and V^* represents an equilibrium solution.

6. Alternative Formulation and Solution Algorithm.

6.1 Alternative Formulations.

Although in theory, some of the available algorithms for solving non linear variational inequalities (see Harker and Pang, 1987) could be directly applied to solve (15), the mathematical difficulty of the formulation makes this approach a non trivial task. Therefore, we propose here a simplification of problem (15) in order to develop a feasible solution algorithm and leave for future research the development of methods that deal directly with (15).

The main difficulty to deal with (15) stems from the presence of nonlinear constraints (12). In fact, the mathematical difficulty of the problem is significantly reduced if we substitute them by the following constraints:

$$v_l^s = x_l^s \cdot f_l \cdot \overline{\omega}^s \cdot V^s, \qquad \forall l \in s, \quad \forall s \in S, \tag{16}$$

In this new formulation, route section flows are assigned to line sections in direct proportion to their frequencies in the attractive set $\overline{A}_s$. Therefore,

congestion is only considered at the level of route sections, and depends on the ratio between flows V^s, $\tilde{V}_s$ and the total capacity of the route sections K_s.

Now, given that line frequencies are not affected by the flows assigned to them, the attractive sets $\overline{A}_s$ will not depend on the level of congestion. Therefore, they can be calculated at the beginning as part of the construction of network $G(\mathcal{N},\mathcal{S})$, which allows us to eliminate the integer variables x_l^s from the mathematical formulation of the problem.

Therefore, the network construction will involve the identification of the route sections and the calculation of the attractive sets $\overline{A}_s$, which is achieved by solving an unrestricted (0,1) hyperbolic programming problem (see, Chriqui and Robillard, 1975; De Cea, 1986).

In our initial formulation, each route section contained all lines that connect its initial and final nodes. A drawback of the simplified formulation is that, given that we select the attractive sets under free flow conditions, "slow" lines can be eliminated from the network representation no matter that they will become attractive when the "fast" ones are full.

Nevertheless, this limitation can be obviated if we introduce the following modification in the calculation of attractive sets: once an attractive set has been calculated for a route section s, the lines in it contained are eliminated and a new attractive set is calculated for the remaining lines; the process is repeated until all lines are introduced in some set. Then, a new network G' is considered, with as many parallel links as attractive sets for each original route section.

Thus, when congestion is low, only the attractive set containing fast lines will be used; nevertheless slow lines will also be used if fast lines become "full". Given the characteristics of transit networks in developing countries, this modification will most probably produce small increases in the number of links to be considered in G, but will make the formulation more rigorous.

Then, the original set Ω can be replaced by a new set Ω' if we substitute the non linear constraints (12) with the simpler linear equations:

$$v_l^s = f_l \cdot \overline{\omega}^s \cdot V^s, \qquad \text{with:} \quad \overline{\omega}^s = \left(\frac{1}{\sum_{l \in s} f_l}\right), \qquad \text{and:} \quad f_s = \sum_{l \in s} f_l, \tag{17}$$

where the V^s are link flows over the new network G'. Thus, (15) becomes:

$$c(V^*) \cdot (V^* - V) \leq 0, \qquad \forall V \in \Omega'. \tag{18}$$

Note than in this case the form of link cost functions is simpler and we can directly use:

$$\varphi_s(V) = \left(\frac{V^s + \widetilde{V}_s}{K_s}\right)^n$$

It is difficult to tell in advance what the consequences of the modifications proposed will be for the solution of real problems, and extensive numerical experiments and comparisons with different models will be required to evaluate them.

6.2 A solution algorithm.

If we consider (15) with constraint set Ω' and defined over the network G', we have a well known transportation assignment problem, for which several efficient algorithms have been proposed. The only peculiarity of the problem is the presence of linear constraints (17), which define variables v_l^s.

Given that the vector cost function $c(V)$ will have in general a non symmetric Jacobian, problem (15) does not have an equivalent optimization formulation. Therefore a solution approach could be using a method to solve directly (15), like the cutting plane algorithm proposed by Nguyen and Dupuis (see Nguyen and Dupuis, 1984) or any other method to solve variational inequalities (see Harker and Pang, 1987). One of the approaches most commonly used to solve non symmetric network assignment problems, because its easy implementation, is

the popularly known "*diagonalization*" method (see Florian 1977, Abdulaal and LeBlanc, 1979). The method corresponds to an iterative approach of the general Jacobi type, for solving non linear equations (see Pang and Chan, 1982). At each iteration the vector cost function $c(V)$ is "*diagonalized*"[4] at the current solution, yielding a symmetric assignment problem. This is mathematically represented by a variational inequality of the form:

$$\widehat{c}(V^*) \cdot (V^* - V) \leq 0, \qquad \forall V \in \Omega', \tag{19}$$

where $\widehat{c}$ has a diagonal Jacobian. Therefore, (19) has an equivalent convex cost optimization problem of the form:

$$\min_{V \in \Omega'} \int_0^V \widehat{c}(x) \cdot dx \tag{20}$$

Then, a sequence of problems (20) are solved until a sufficiently stable solution is obtained. Several efficient algorithms exist for solving problem (20), of which the most popular is the Frank-Wolfe descend method (LeBlanc et al, 1975).

The solution approach to problem (18) can be summarized as follows:

Step 0: (initialization). Find an initial feasible solution $(\overline{V}, \overline{v})$.

Step 1: Diagonalize $c(V)$ on $(\overline{V}, \overline{v})$.

Step 2: Solve problem (20) to get $(\widehat{V}, \widehat{v})$

Step 3: Stopping test. If $(\overline{V}, \overline{v})$ and $(\widehat{V}, \widehat{v})$ are sufficiently close ,**stop**; otherwise:

Step 4: Make: $(\overline{V}, \overline{v}) \leftarrow (\widehat{V}, \widehat{v})$, and return to *Step 1*.

Let us briefly turn our attention to Step 2. If we use the Frank-Wolfe algorithm to solve (20) the only difference with an standard traffic assignment problem comes from the presence of linear equations (17), which involve variables v_l^*.

[4] A non linear mapping $F(x) : \Re^n \rightarrow \Re^n$ is "*diagonalized*", if each of its components $f_i(x)$ is made only a function of x_i (see Ortega and Reinboldt, 1970).

However, note that as a consequence of the diagonalization those variables do not appear any more in the objective function, which depends only on variables V^s.

Then, constraints (17) are completely separable within each diagonalization iteration and Step 2 can be solved as an standard traffic assignment problem in variables V^s, over the network G', at the end of which variables v_l^s can be calculated using equation (17).

Convergence of the algorithm will be based on the usual monotonicity conditions required for $c(V)$ (see Florian and Spiess, 1982). Nevertheless, it is important to remind that such conditions are sufficient but not necessary and that "diagonalization" algorithms have shown good convergence properties in practice even when monotonicity is not satisfied.

Note that if we eliminate from consideration the congestion term φ_s in cost function (3), we obtain an algorithm for solving the uncongested case, which will yield the same solution as Chriqui's method, as was mentioned at the beginning of section 3. In such case it is obviously unnecessary to create parallel links in G', for each route section. The algorithm will reduce to calculate shortest paths over G', and then assigning to them the O-D transit flows, which in turn will be split between transit lines in proportion to their relative frequencies on the route sections belonging to the shortest paths.

REFERENCES

M. ABDULAAL and L.J. LEBLANC, "Methods for Combining Modal Split and Equilibrium Assignment Models," *Trans. Sci.* **13**, 292–314 (1979).

C. CHRIQUI, " Réseaux de Transport en Commun: les problèmes de cheminement et d'accès," Publication No. 11, Centre de Recherche sur les Transports, Université de Montréal (1974).

C. CHRIQUI and P. ROBILLARD, "Common Bus Lines," *Trans. Sci.* **9**, 115–121 (1975).

J. DE CEA, " Rutas y Estrategias Optimas en Modelos de Asignacion a Redes de Transporte Público," IV Congreso Panamericano de Ingeniería de Tránsito y Transporte, Santiago, Chile, Diciembre (1986).

J. DE CEA, L. ZUBIETA, M. FLORIAN, P. BUNSTER, "Optimal Strategies and Optimal Routes in Public Transit Assignment Models," Publication No. 555, Centre de Recherche sur les Transports, Université de Montréal (1987).

J.E. FERNANDEZ and F. BRAVO, "Modelación de Sistemas de Transporte Público Urbano", III CLAIO, 18–22 Agosto. Santiago, Chili (1986).

M. FLORIAN, "A Traffic Equilibrium Model of Travel by Car and Public Transit Modes," *Trans. Sci.* **8**, 166-179 (1977).

M. FLORIAN and H. SPIESS, "The Convergence of Diagonalization Algorithms for Asymmetric Network Equilibrium Problems," *Trans. Res.* **16B** 447–483 (1982).

M. FLORIAN and H. SPIESS, "On Binary Mode Choice/Assignment Models," *Trans. Sci.* **17–1**, 32–47 (1983).

P. T. HARKER and J-S. PANG, "Finite-Dimensional Variational Inequality and Nonlinear Complementarity Problems: A Survey of Theory, Algorithms and Applications," Department of Decision Sciences, The Wharton School, University of Pennsylvania, 87- 12-06, (1987).

A. LAST and S.E. LEAK, "TRANSEPT: A Bus Model", *Traff. Eng. and C.* (1976).

L.J. LEBLANC, E.K. MORLOK and W.P. PIERSKALLA, "An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment Problem,"

Trans. Res. **5** 309–318 (1975).

S. NGUYEN and C. DUPUIS, "An Efficient Method for Computing Traffic Equilibria in Networks with Asymmetric Transportation Costs," *Trans. Sci.* **18**, 185–202 (1984).

S. NGUYEN and S. PALLOTINO, "Equilibrium Traffic Assignment for Large Scale Transit Networks", to appear in the European Journal of O. R. , (1988).

J.M. ORTEGA and W.C. REINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

J.M. PANG and D. CHAN, "Iterative Methods for Variational and Complementarity Problems," *Math. Prog.* **24**, 284–313 (1982).

H. SPIESS, "On Optimal Route Choice Strategies in Transit Networks," Pub. 286, Centre de Recherche sur les Transports, Université de Montréal (1983).

H. SPIESS and M. FLORIAN, "Optimal Strategies: A New Assignment Model for Transit Networks", Centre de Recherche sur les Transports, Université de Montréal (1988).

APPENDIX

NUMERICAL EXAMPLE

1. Introduction

In order to illustrate the operation of the algorithm proposed in section 6.2 of the paper we develop a numerical example using the following transit network, taken from Spiess and Florian, (1988):

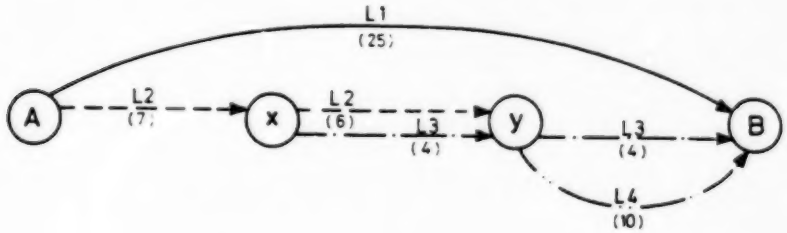


FIG. A1. Example transit network.

"Take the network represented in Figure A1. composed of one O–D pair **A–B** and two transfer nodes, **X** and **Y**.

There are four transit lines serving such network:

L1. An express line going directly from **A** to **B**, with a service frequency (f_l) of 10 buses per hour. In vehicle travel time is 25 minutes.

L2. Line going from **A** to **Y** with an intermediate stop at **X**. Travel times on each line segment are 7 and 6 minutes respectively and the service frequency is 10.

L3. Goes from **A** to **B** with intermediate stop at **Y**. Travel times are of 4 minutes on each segment and frequency is 4.

L4. Runs between **Y** and **B** with a frequency of 20 and travel time of 10 minutes.

Therefore average waiting times on each line are of 6, 6, 15 and 3 minutes respectively."

The modified transit network $G = (\mathcal{N}, \mathcal{S})$ is represented in Figure A2. and the basic data related with each link (route section) in G, is given in Table 1. below.

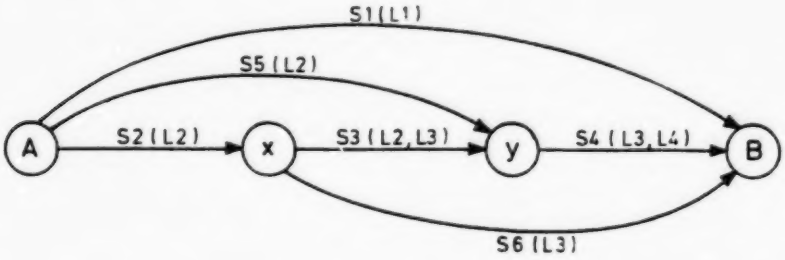


FIG. A2. Modified transit network $G = (\mathcal{N}, \mathcal{S})$.

BASIC DATA		G *NETWORK Links*					
		$S1$	$S2$	$S3$	$S4$	$S5$	$S6$
$\bar{c}_s$	(min)	25	7	5.4*	9.0*	13	8
(α/f_s)	(min)	6	6	4.3	2.5	6	15
$\tilde{c}_s$	(min)	31	13	9.7	11.5	9	23
K_s	(pass/hr)	100	100	140	240	100	40

*(expected travel time values)

Table A1. Basic link data for modified network G.

Notice that for links $S3$ and $S4$ the in-vehicle travel times are equal to the expected values experienced by passengers using those sections. Given that we are assuming that the probability of taking each line is proportional to its frequency, each line travel time value is weighted by the corresponding relative frequency. The value of parameter α is taken equal to 1, and total capacity for each route section (K_s)is calculated assuming 10 passengers transit vehicles.

2. Volume-Delay Functions

In order to simplify the calculations the following link cost functions, linear in the link flows, are assumed:

$$
\begin{aligned}
c_1 &= 31 + \left(\frac{V^1}{10}\right), & c_2 &= 13 + \left(\frac{V^2+V^5}{10}\right), \\
c_3 &= 9.7 + \left(\frac{V^3+V^5+V^6}{14}\right), & c_4 &= 11.5 + \left(\frac{V^4+V^6}{24}\right), \\
c_5 &= 19 + \left(\frac{V^5+V^2}{10}\right), & c_6 &= 23 + \left(\frac{V^6+V^3+v_2^3}{4}\right).
\end{aligned}
\qquad (A1)
$$

Then, the vector cost function $c(V)$ has the following Jacobian:

$$
\left(\frac{\partial c}{\partial V}\right) = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}
\begin{array}{c}
\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\
\left(\begin{array}{cccccc}
1/10 & 0 & 0 & 0 & 0 & 0 \\
0 & 1/10 & 0 & 0 & 1/10 & 0 \\
0 & 0 & 1/14 & 0 & \mathbf{1/14} & \mathbf{1/14} \\
0 & 0 & 0 & 1/24 & 0 & \mathbf{1/24} \\
0 & 1/10 & \mathbf{0} & 0 & 1/10 & 0 \\
0 & 0 & \mathbf{1/4} & \mathbf{0} & 0 & 1/4
\end{array}\right)
\end{array}
$$

We can see that the Jacobian is non-symmetric and non monotone, given that the value of $\det(B + B^T) = -3.6 \cdot 10^{-6}$, (where B represents the Jacobian).

In order to show the behavior of the algorithm under different transit operating conditions, the following three cases are examined: uncongested, mild congestion, and high congestion. Nevertheless, we make notice that the linear form assumed (for simplicity) for the cost functions is not the most appropriate to handle high congestion, given that it does not allows to impose a high penalty on those network links that are overloaded. For that, a convex increasing function should be used. Based on the experience we have with car network assignment algorithms a power function with n between 4 and 6 should be appropriate.

3. Uncongested Case

For this case a total travel demand T_{AB} equal to 1 is assumed. This makes the congestion term (dependent on flows V^i) to have a negligible influence on travel costs.

Then, the minimum path between **A–B**, on network G, is: **(A,S5,S4,B)** and given that in this case all the flow is allocated to the minimum path, the following flow vector is obtained: $V = (0,0,0,1,1,0)$. If link flows are now assigned to transit lines in proportion to their relative frequencies on the corresponding route section (using equation (16) of the paper),the following line segment flows are obtained (other flows are zero):

$$v_2^5 = 1, \quad v_3^4 = 0.17 \quad \text{and} \quad v_4^4 = 0.83$$

The solution is graphically represented in Figure A3, and can be compared with that provided by using Spiess' algorithm, represented in Figure A4; notice that Spiess' algorithm spreads the flows over more paths, using line L1. This is due to the use of the concept of "optimum strategy" (see Spiess and Florian, 1988).

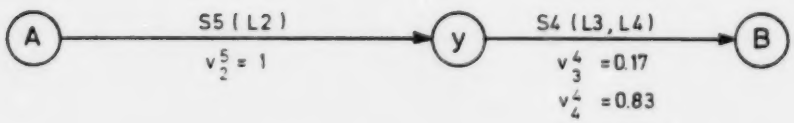


FIG. A3. Uncongested case solution.

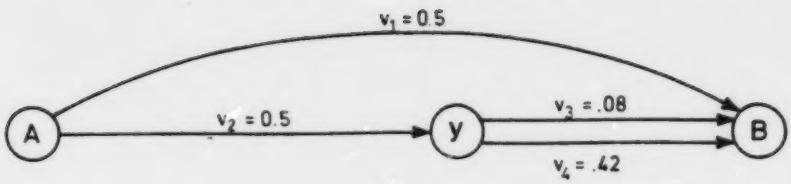


FIG. A4. Spiess solution, (optimum strategy).

4. Mild congestion

In this case a travel demand $T_{AB} = 100$ is assumed. Thus, the contribution of the congestion term to the travel cost will start to have some significance though still less important than the fix cost.

In order to solve the non-symmetric assignment problem over G, the Jacobi (diagonalization) method proposed in section 6.2 of the paper will be used. Taking the solution to the uncongested case as initial solution:

$$V_0 = (0, 0, 0, 100, 100, 0),$$

the following "diagonal" costs are obtained for the first Jacobi iteration:

$$\begin{aligned} c_1 &= 31 + \left(\frac{V^1}{10}\right), & c_2 &= 23 + \left(\frac{V^2}{10}\right), \\ c_3 &= 16.8 + \left(\frac{V^3}{14}\right), & c_4 &= 11.5 + \left(\frac{V^4}{24}\right), \\ c_5 &= 19 + \left(\frac{V^5}{10}\right), & c_6 &= 23 + \left(\frac{V^6}{4}\right). \end{aligned} \qquad (A2)$$

Table A2 shows the Frank-Wolfe iterations necessary to solve the diagonal problem.

$F-W$	$FLOWS^{(1)}$	$FIRST\ \ STEP$		$2^{nd}\ \ STEP$
(IS)	(0,0,0,100,100,0)	MP^2	$AUX.FLOWS$	λ^3
1		$(A, S1, B)$	(100,0,0,0,0,0)	0.570
2	(57,0,0,43,43,0)	$(A, S5, S4, B)$	(0,0,0,100,100,0)	0.008
3	(57,0,0,43,43,0)	$STOP$		

[1] (Current feasible flows)
[2] (Minimum path for the current feasible flows)
[3] (Line search solution)

Table A2. Frank-Wolf solution to the diagonal assignment on G.

Only two iterations are needed to arrive to the solution:

$$V^* = (57, 0, 0, 43, 43, 0),$$

which turns out to be also the equilibrium solution to the global problem. This is due to the fact that the Jacobian is diagonal if we only take into account the travel cost functions corresponding to links with flow.

$$\left(\frac{\partial c}{\partial V}\right)^* = \begin{array}{c} \\ 1 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 4 & 5 \end{array} \\ \begin{pmatrix} 1/10 & 0 & 0 \\ 0 & 1/24 & 0 \\ 0 & 0 & 1/1 \end{pmatrix} \end{array}$$

If link flows are now assigned to the corresponding transit lines in proportion to their relative frequencies, the following solution is obtained (all other flows are zero):

$$v_1^1 = 57, \quad v_2^5 = 43, \quad v_3^4 = 7, \qquad v_4^4 = 36.$$

The solution is represented in Figure A5. As we can see the structure of the solution is the same as that obtained using the Spiess algorithm, although the flows are not the same because in our case they depend on the form of the congestion terms used for the cost functions.

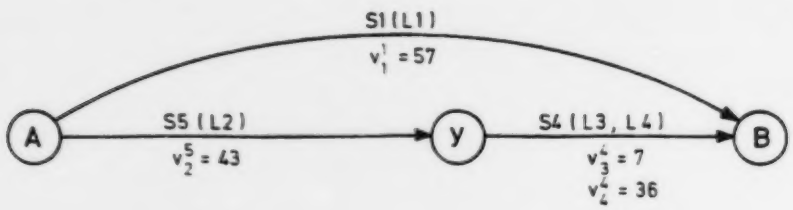


FIG. A5. Mild congestion solution.

5. High Congestion

In order to represent a case where congestion be important, the value of the congestion parameter β is increased to 20 and the demand T_{AB} is set equal to 240. The solution obtained in this case is:

$$V^* = (140, 6, 0, 94, 94, 6),$$

which is represented in Figure A6 where the corresponding line section flows are also indicated. As we can see, now the path $(A, S2, S6, B)$ is also used due to increased congestion on the other paths. It can also be observed that line 1 is overloaded in 40 units because the linear cost functions used do not impose a high enough penalty to overloading.

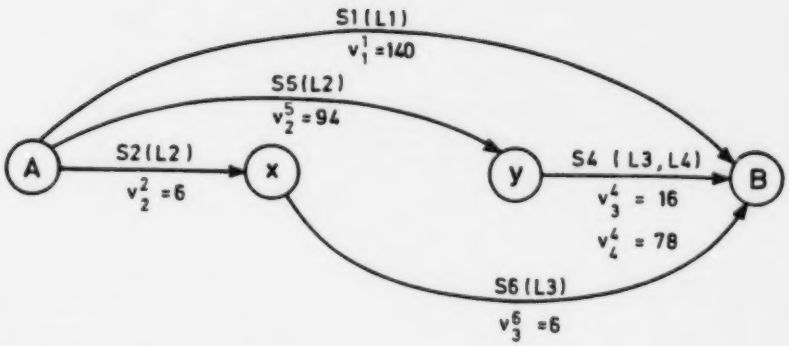


FIG. A6. High congestion solution.